Inter-industry Labor Flows

Frank Neffke, Anne Otto and Antje Weyh
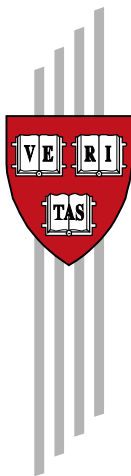
Working Papers

Center for International Development
at Harvard University

Inter-industry labor flows

Authors

Frank Neffke[1], Anne Otto[2], Antje Weyh[2]

[1] Harvard Kennedy School, Center for International Development,

[2] Institute for Employment Research (IAB)

Abstract

Labor flows across industries reallocate resources and diffuse knowledge among economic activities. However, surprisingly little is known about the structure of such inter-industry flows. How freely do workers switch jobs among industries? Between which pairs of industries do we observe such switches? Do different types of workers have different transition matrices? Do these matrices change over time? Using German social security data, we generate stylized facts about inter-industry labor mobility and explore its consequences. We find that workers switch industries along tight paths that link industries in a sparse network. This labor-flow network is relatively stable over time, similar for workers in different occupations and wage categories and independent of whether workers move locally or over larger distances. When using these networks to construct inter-industry relatedness measures they prove better predictors of local industry growth rates than co-location or input-based alternatives. However, because industries that exchange much labor typically do not have correlated growth paths, the sparseness of the labor-flow network does not necessarily prevent a smooth reallocation of workers from shrinking to growing industries. To facilitate future research, the inter-industry relatedness matrices we develop are made available as an online appendix to this paper.

1. Introduction

How mobile are workers across industries? Which industries can easily exchange labor? Such questions matter, because on the one hand, shocks to an economy's industrial structure require the transfer of productive capacity, and thus of workers, from shrinking to growing industries. On the other hand, labor mobility transfers the knowhow of workers across firms, industries and locations. This makes labor mobility an important factor in organizational learning (March, 1991; Simon, 1991) and regional and national growth (Saxenian, 2006). However, in spite of their importance, surprisingly little is known about the structure of inter-industry labor flows. As a result, we have incomplete answers to a number of important questions: How large is the set of industries a worker chooses from when he or she changes jobs? Do different types of workers switch among the same industries, i.e., do they follow the same industrial transition matrix? Does the network this matrix represents change over time? Is this labor-flow network capable of identifying sets of industries that are related in terms of their human capital requirements? And, finally, to what extent do the mobility constraints expressed in this network prevent an economy from reallocating labor from shrinking to growing industries? To answer these questions, we develop a number of statistical tools for analyzing inter-industry labor flows at a high level of detail and use these tools to formulate a number of stylized facts that emerge from the inter-industry labor flows in the German economy.

These stylized facts speak to two different literatures and we organize the paper accordingly. Firstly, our paper relates to the literature human capital specificities in the field of labor economics and in particular to the work on workers switching jobs. Our conceptual starting point is that, if jobs require industry-specific human capital, the mobility of workers across industries will be constrained, and inter-industry labor flows will be shaped by those constraints. Recent empirical work on job switching suggests, however, that such constraints are not particularly severe. For instance, workers in the United States

change 1-digit industries at rates of between 13% (Kambourov and Manovskii, 2008) and 20% (Parrado et al., 2007) a year. However, this research fails to take into consideration *which* industries exchange labor, implicitly assuming that all industries are equidistant from one another in terms of their human capital requirements. If this assumption does not hold, evidence drawn from industry-switching rates could be misleading. This indeed seems to be the case. Drawing on a data set that covers over 80% of the German workforce between 1999 and 2008, we find that, although workers do often switch industries even at a very high level of aggregation (stylized fact 1), labor flows are still highly structured (stylized fact 2). In particular, 50% of the workers leaving an industry are subsequently employed in sets of industries (each set being specific to the industry that workers leave) that together represent less than 4% of total German employment.

Secondly, our paper relates to the recent use of inter-industry labor flows as a way of measuring human capital similarities or *skill relatedness* among industries (e.g., Greenstone et al., 2010; Dauth, 2010; Neffke and Henning, 2013). The rationale behind this use of labor flows is that labor flows arise among industries with similar human capital requirements. In this light, the abovementioned observation that labor flows are highly structured and limited to a small number of industry pairs would imply that most industries have quite incompatible skill sets. Accordingly, the underlying anatomy of the labor market is structured as a skill-relatedness network in which industries are sparsely connected to one another through labor flows.

When exploring these skill-relatedness structures, we find that they are remarkably stable over the-period studied (stylized fact 3). Moreover, job switches taking place within a region follow a very similar skill-relatedness structure as job switches over longer distances (stylized fact 4), ruling out that skill-relatedness measures simply reflect industries' geographical co-location patterns. Finally, by calculating skill-relatedness matrices for different labor-market segments, we find that the implied skill-relatedness

networks are remarkably similar for workers in different occupations, at different wage levels and in different parts of the country (i.e., former East and West Germany, stylized fact 5).

Next, we use the skill-relatedness measure to address a growing field of research that has employed a variety of industry relatedness measures to predict growth of industries in countries (Hidalgo et al., 2007) and regions (e.g., Porter, 2003; Neffke et al., 2011; Delgado et al., 2010). These papers find that industries grow faster in regions where they are surrounded by many related economic activity. A particularly prominent approach in this literature derives relatedness from the geographical co-location of industries (e.g., Porter, 2003; Hidalgo et al. 2007). However, in a direct comparison, our labor-flow based measure typically outperforms such co-location-based measures, as well as input-output-based measures, in predicting local industry entry and growth rates (stylized fact 6). Although, these predictions would suggest that regions with many skill-related industries are well-positioned for future growth, because labor is constrained to move among skill-related industries, such regions might be vulnerable if skill-related industries are typically exposed to correlated shocks. However, we find that this is not the case: skill-related industries have uncorrelated growth patterns. Consequently, the skill-relatedness network does not seem to represent a major impediment to the reallocation of labor from shrinking to growing industries (stylized fact 7).

Overall, the stylized facts we find suggest that inter-industry labor flows and the skill relatedness expressed in them are interesting objects of study. Therefore, to facilitate further research, we make the skill-relatedness matrices used in this paper, together with equivalent ones for each of the industry classification systems that have been in use in Germany between 1975 and 2014, available in an online appendix to this paper.

The paper is structured as follows. In section 2, we discuss the literature on human capital specificities and job switches and the literature on inter-industry relatedness measures. Section 3 describes the data.

In section 4, we develop a number of statistical tools to analyze labor flow networks which are described in greater detail in the appendix to this paper. Furthermore, section 4 presents the stylized facts we uncover using these tools. Section 5 discusses future research and concludes.

2. Labor flows and inter-industry relatedness

Human capital and skills are pivotal inputs in today's production processes. Accordingly, in management science, a firm's workforce is regarded as an exceedingly important competitive asset (Porter, 1987; Grant, 1996; Grant and Spender, 1996). Moreover, today's workforces are highly specialized: individual workers often invest heavily in education and training to acquire specific skills that allow them to carry out tasks that range from engineering or financial management to construction work or food preparation. Because workers specialize, their human capital is often held to be specific to the firm where they work (Becker, 1964), to an industry (Neal, 1995; Parent, 2000; Sullivan, 2010) and to the tasks performed (Poletaev and Robinson, 2008; Gathmann and Schoenberg, 2010). However, there is considerable debate about which of these dimensions of skill specificity dominate. For instance, Kambourov and Manovskii (2009) study the value of occupation tenure and industry tenure and report that only the former is rewarded with higher wages, casting doubt on the existence of industry specificities in human capital. In contrast, Sullivan (2010) shows that industry tenure sometimes reaps high rewards, but that the extent to which this happens depends on the occupation.

We propose a different approach to answer the question of whether human capital has an industry-specific component. The starting point is that valuable information on this question is contained in the exact job switching patterns of workers. After all, whenever the old and the new job require different skills, job switches will render some human capital redundant. To avoid this kind of human capital depreciation, workers will predominantly switch to jobs that allow them to reuse their skills. This

suggests that overlap in industries' human capital requirements, or, more accurately, absence of such overlap, should constrain inter-industry labor movements. Therefore, job switches between different industries not only reveal whether there is an industry component to human capital, but also which industries hire workforces with similar skills and know-how.

The insight that labor flows contain information on the relatedness among industries in terms of the human capital they employ has been leveraged in a number of recent papers. For instance, Greenstone et al. (2010) use labor transitions in the Current Population Survey (CPS) to show that large-plant openings create spillovers to local firms, but in particular to firms in industries that are related to the new plant's industry in terms of labor transitions. Similarly, inter-industry relatedness measures based on labor flows have been used in other studies in economic geography (Dauth, 2010; Timmermans and Boschma, 2014), trade (Kaplan et al., 2011), strategic management (Neffke and Henning, 2013) and entrepreneurship research (Costa and Baptista, 2011). To our knowledge, however, there has hitherto not been any detailed investigation of the structure of inter-industry labor flows to shed light on the legitimacy of such labor-flow based relatedness measures.

3. Data

Our data are constructed from Germany's social security records as compiled in the Historic Employment and Establishment Statistics (HES) database.[1] The HES offers a large set of demographic and employment characteristics, such as a worker's daily wage,[2] occupation, work status (i.e.,

---

[1] Bender et al. (2000) provide a detailed description of this database.

[2] We deflate wages to 2005 EUR.

apprentice, part-time worker, full-time worker), gender and age. Furthermore, the industry[3] and location of each individual's work establishment are known.

Because of changes in the industry classification system, we confine our analyses to the years 1999 to 2008. Furthermore, we limit the analyses to full-time employees aged 18 to 65. We exclude apprentices and volunteers because they are still investing in education to acquire skills. Because wage information is left-censored due to upper limits to social security contributions, we impute wages whenever they exceed these upper bounds following the method of Gartner (2005). Due to changes in the industry-classification system (see for a more detailed description Appendix A), we confine our analyses to the years 1999 to 2008. This results in a final data set on, on average, about 20 million workers a year.

Definition of labor flows

We use the HES to construct inter-industry labor flows. Labor flows arise when workers switch establishments from one year to the next. Unfortunately however, establishment identifiers in the HES are not perfectly reliable. New identifiers are assigned to establishments in case of spin-offs, mergers, break-ups or recodings. Hethey and Schmieder (2010) find that for only 35% to 40% of all establishments with over three employees a new (or a disappearing) establishment identifier can be interpreted unambiguously as an entry (or as an exit). In the other cases, workers move in larger blocks from one establishment to another. To avoid that such spurious identifier changes contaminate our labor flow measurements, we remove 531,000 job switches (27.5%) from a total of the 1.8 million yearly job switches.[4]

Labor-market segments

---

[3] We drop workers employed through employment agencies, because we don't have information on the actual industry or region in which these individuals work.

[4] See Appendix B for a more detailed description of the identification and elimination of spurious job switches.

In the empirical section of this paper, we decompose the labor market into different segments. For each labor-market segment we create separate flows, based on the segment from which the flow originates. We introduce three kinds of labor market segmentations, one based on the geography of the flows, one that captures workers' skills and one that distinguish between an eastern and a western German labor market.

The first labor segmentation is motivated by the potential concern that the structure of inter-industry labor flows is governed by the availability of local jobs. In that case, inter-industry labor flows could simply be an expression of the co-location patterns of industries. We look into this by comparing job switches that occur over short distances to long-distance switches, i.e., switches for which the old and the new job are at least 100 kilometers apart.[5] Second, to proxy workers' human capital *levels*, we segment the labor market into workers who earn below and those who earn above the median wage in their industries. To explore whether the *type* of human capital matters, we next separately analyze workers in eight broad occupational groups that are associated with different broad sets of tasks. In particular, we distinguish among managers, sales-related employees, accountants, information technology (IT) workers, office clerks, cleaners, security personnel and other workers (see Appendix C). These occupations were chosen because of their relative ubiquity across a variety of industries. Third, we investigate whether there are regional differences in labor-flow patterns by splitting the German labor market along the former border between East and West Germany.[6]

---

[5] Distances between old and new jobs represent road distances between the centroids of the districts (*Kreise*) in which the corresponding establishments are registered. Given that only five percent of German employees commute over distances greater than 50 km (Winkelmann, 2010), we postulate that distances of over 100 km typically require a worker to relocate.

[6] To ensure that our results are not driven by the extensive outmigration from the East after Germany's reunification, we exclude flows between East and West Germany in these analyses.

4. Results

Below, we derive a number of stylized facts using the data described in the previous section. We first describe the general structure of the labor flows in terms of the amount of job switches and the degree to which labor flows concentrate in relatively few industry pairs. Next, we analyze the general structure underlying these flows by plotting the skill-relatedness network and comparing the skill-relatedness matrices for different labor-market segments. Finally, we turn to the question of how skill relatedness affects local labor markets by estimating local industry-growth regressions and determining the extent to which the limited mobility of workers across industries could potentially hinder an efficient reallocation of workers from shrinking to growing industries. These analyses draw on a number of statistical tools and indicators for which the intuition is described in the main text, whereas details and derivations are described in the appendix to this paper.

Cross-industry labor-flow patterns

At their coarsest level of aggregation, industries in the European NACE Revision 1.1 classification are divided into sections. Sections consist of several sub-sections, which themselves are made up of 2-digit industries. These 2-digit industries are further subdivided into 3-, 4-, and 5-digit industries. Table 1 summarizes average yearly labor flows across industries at these different levels of aggregation. The first column of Table 1 reports figures for Germany as a whole. Of all workers who change jobs, 73.4% change industries at the 5-digit level, the most disaggregated level available. In line with Kambourov and Manovskii's (2008) finding that much industry switching takes place across highly aggregated industries, we find that 58.7% of these 5-digit industry switchers switch industries at the section level, the most aggregated industry grouping in the NACE 1.1 classification.

Distinguishing flows by their labor-market segment of origin, we find that workers in the high-income segment switch industries less often and undertake less drastic switches than low-wage workers do. For

instance, 8.1% of high-wage workers who switch industries stay in their 4-digit industries (column 2), against only 6.2% for low-wage workers (column 3). Workers in eastern Germany switch industries somewhat less than their colleagues in the west (columns 4 and 5). Columns 6 and 7 show that jobs switches over distances below 100 km ("local" labor flows) display patterns that are very similar to those that involve distances of over 100 km ("long-distance" labor flows). Differences by occupation are larger.[7] Workers in lower-skilled occupations such as cleaning (78.8%) and security (73.9%) cross section borders much more often than workers in higher-skilled occupations, such as management (58.4%), accountancy (58.4%) and IT (58.5%) jobs, who tend to switch to industries that are classified more closely to their old jobs. This suggests that moving to a radically different industry is less attractive when human capital requirements are higher. However, in spite of these differences among labor-market segments, overall, the following stylized fact emerges from Table 1:

Stylized fact 1: Workers frequently cross industry boundaries, even at the highest level of aggregation in the industry classification system.

TABLE 1 (FLOWS) ABOUT HERE

Flexibility

The fact that workers tend to cross boundaries even between highly aggregated sectors does not necessarily mean that labor flows are unstructured. Indeed, 3.3% of all possible pairs of 5-digit industries account for 80% of overall job switches and in 56% of all pairs we do not observe any job switches whatsoever in the nine-year period we study. How random then are labor flows? To answer this

---

[7] Workers in security and cleaning jobs seem to switch industries less often than other workers. However, this is misleading because cleaning and security employees work predominantly in either the cleaning or the security industry. As a consequence, most job switches occur within these two 5-digit industries. Once cleaners and security guards cross industry boundaries, their labor flows look much less structured than the ones in other occupations.

question, we calculate for each industry the minimal set of industries that together absorb 50% of the industry's labor outflow. Let F_{ij}^S be the labor flow originating in segment s from industry i to industry j. Furthermore, let α_{ij}^S be:

$$\alpha_{ij}^S = \frac{F_{ij}^S}{N_j^S}$$

where N_j^S is the number of workers in industry j in segment s. α_{ij}^S can be thought of as the rate per employee in industry j at which the industry absorbs workers from industry i. Next, for each industry of origin, we sort destination industries in descending order of this ratio: $\alpha_{i1}^S > \alpha_{i2}^S > \cdots > \alpha_{in}^S$, where n is the number of industries in the economy. Let $F_{i.}^S = \sum_j F_{ij}^S$ be the total labor outflow in a segment s from industry i and $N^S = \sum_i N_i^S$ the total employment in this segment. We define $FLEX_i^S(q)$ as:

$$FLEX_i^S(q) = \sum_{j=1}^{k} \frac{N_j^S}{N^S} \quad \text{with } k = \operatorname{argmin}_{k'} \left(q < \sum_{j=1}^{k'} \frac{F_{ij}^S}{F_{i.}^S} \right)$$

In other words, $FLEX_i^S(q)$ gives the minimum employment share of a set of 5-digit industries that together absorb at least $q\%$ of all workers who leave industry i in segment s.[8] To determine the flexibility of workers in an entire labor-market segment, we calculate the outflow-weighted average of $FLEX_i^S(q)$ across all industries of origin, i:

$$FLEX^S(q) = \sum_i \frac{F_{i.}^S}{F_{..}^S} FLEX_i^S(q)$$

where $F_{..}^S = \sum_i F_{i.}^S$ represents the total inter-industry labor flows that originate from labor-market segment s.

$FLEX^S(q)$ can be calculated both with and without within-industry labor flows, that is with or without workers who change establishments, but not industries. Figure 1 depicts $FLEX^G(q)$ against q where G

---

[8] Where necessary, we interpolate $\sum_{j=1}^{k} \frac{N_j}{N}$ between k and $k+1$.

is the labor-market segment that contains the entire German labor market, averaged over the period 1999 to 2007. $FLEX^G(q)$ is highly convex, meaning that the bulk of labor outflows are absorbed by a small part of the economy. Because some of the curvature of $FLEX^G(q)$ would also have occurred if flows simply followed employment, we create a benchmark based on simulated flows that reflects the value of $FLEX^G(q)$ if workers had moved randomly across industries. In this simulated benchmark, industry switchers end up in industry j with probability $p_j^S = \frac{N_j^S}{N^S}$, $\sum_j p_j^S = 1$, regardless of their industry of origin. The resulting null-model curve is depicted as a dotted line.

FIGURE 1 (FLEX) ABOUT HERE

Table 2 compares the flexibility of workers in different labor-market segments by providing $FLEX(0.50)$ values for each segment excluding (column 1) and including (column 2) within-5-digit-industry flows. The number in parentheses shows the corresponding random benchmark.

TABLE 2 (FLEX) about here

In general, labor flows are tightly structured, especially if we include within-industry flows.[9] For Germany as a whole, 50% of all workers move to sets of industries[10] that represent only 3.4% (5.9% if within-industry flows are excluded) of the economy. High-wage workers, with $FLEX(0.50)$ equal to 2.2% (3.9%), are much less flexible than low-wage workers, who exhibit a $FLEX(0.50)$ value of 4.0% (6.4%). Given that low- and high-wage groups have virtually identical null-model predictions, their estimates can be easily compared. For other labor-market segments, null-model predictions vary widely,

---

[9] The exception is occupational segments that are dominated by a single industry (as in cleaning and security jobs), where most of the flows take place among firms within that 5-digit industry. Here including the within-industry flows substantially decreases the estimated flexibility.

[10] Note that these sets differ by industry, that is, each industry typically has a different set of industries that absorbs most of its worker outflow.

which complicates comparisons across groups. However, given that $FLEX(0.50)$ values are well below their null-model predictions in all segments, we arrive at the following stylized fact:

> **Stylized fact 2:** Labor flows are channeled through tight paths: most workers are absorbed by sets of industries that represent only a small fraction of total employment in the economy.

The skill-relatedness structure of labor-flow matrices

So far, we have documented patterns in the raw labor flows. However, the size of labor flow will depend on the sizes and flow rates (the fraction of employees switching jobs) of the industries involved. To isolate the structure underlying inter-industry labor flows, we calculate the ratio between the observed volume of labor flows, and the one that would be expected from industries' flow rates. If workers switched industries with probabilities proportional to the total outflow of the industry of origin, $F_{i.}^{S}$, and the total inflow into the destination industry, $F_{.j}^{S}$, the expected labor flows between i and j is given by $\hat{F}_{ij}^{S} = \frac{F_{i.}^{S} F_{.j}^{S}}{F_{..}^{S}}$ and the ratio of observed to expected flows by:[11]

$$R_{ij}^{S} = \frac{F_{ij}^{S} F_{..}^{S}}{F_{i.}^{S} F_{.j}^{S}} \tag{1}$$

---

[11] Alternatively, we can derive a baseline expectation using the size of industries. That is, if we assume that workers switch from one industry to another industry with probabilities that are proportional to these industries' sizes, then the expected labor flow from industry i to j is $N^{S} \frac{F_{..}^{S}}{N^{S}} \frac{N_{i}^{S}}{N^{S}} \frac{N_{j}^{S}}{N^{S} - N_{i}^{S}}$. Relatedness is now calculated as:

$$R_{ij}^{s} = \frac{F_{ij}^{S}(N^{s} - N_{i}^{s})}{F_{..}^{S} N_{i}^{s} N_{j}^{s}}$$

Equation (1) yields R_{ij} values that have slightly higher year-on-year correlations, suggesting a somewhat higher consistency. In practice, however, both baselines give virtually the same results, with rank correlations of typically around 0.98. An intermediate solution is provided in Neffke and Henning (2013), who use regression analysis to predict expected flows from general industry level characteristics. Also this method yields very similar results.

Values for R_{ij}^S from 1 to infinity indicate that labor flows are in excess of our random benchmark. Values between 0 and 1 indicate that labor flows are below their random benchmark. Because the distribution of R_{ij}^S is strongly right-skew, we transform R_{ij}^S as follows:

$$\bar{R}_{ij}^S = \frac{R_{ij}^S - 1}{R_{ij}^S + 1}$$

which maps R_{ij}^S values between 0 and 1 onto the interval $[-1,0]$ and values from 1 to infinity onto the interval $[0,1)$. As a result, $\bar{R}_{ij}^S$ is symmetrically distributed around zero, where, for instance, flows twice or half the random baseline are at equal distance from, yet opposite sides, of zero.[12] In line with the notion that inter-industry labor flows express skill similarities, we refer to $\bar{R}_{ij}^S$ as the *skill relatedness* of i to j in segment s and call i and j *skill related* if $\bar{R}_{ij}^S > 0$.

Figure 2a and 2b show two visualizations of the $\bar{R}$-matrix for the German labor market as a whole. Figure 2a shows a heat map for skill-relatedness estimates among all possible 5-digit industry pairs, with rows and columns sorted by an average-linkages hierarchical clustering algorithm. The dark squares along the figure's diagonal indicate that the matrix exhibits a fair degree of clustering, i.e., sets of skill-related industries are connected among each other. However, there are also links across these clusters, as evidenced by the various dark off-diagonal areas.

To get an impression of which industries are connected, Figure 2b shows the network spanned by just the top 651[13] values in the $\bar{R}$-matrix, instead of depicting the entire skill-relatedness matrix. In this figure, nodes represent 3-digit[14] industries (colored according to the sections – the NACE classification's highest level of aggregation – to which they belong). The size of a node represents the corresponding

---

[12] In particular, observed flows that exceed expected flows by a factor a translate into $\bar{R}_{ij}^S = \frac{a-1}{a+1}$, whereas the opposite – expected flows exceeding observed flows by a factor a – yields $\bar{R}_{ij} = \frac{1/a - 1}{1/a + 1} = \frac{1-a}{1+a} = -\frac{a-1}{a+1}$.

[13] We display three times as many links as nodes, which, as a rule of thumb, yields networks that are not overly cluttered.

[14] Aggregating industries at the 3-digit level allows us to label the majority of nodes.

industry's average employment in the period we study. The layout of the network is based on an algorithm that aims at grouping closely related industries together such that nodes that cluster in the graph generally correspond to sets of skill-related industries.

FIGURE 2b (IND SPACE) about here

Industries tend to cluster by section. For instance, we find a cluster of dark-blue textiles and leather industries (center-left), a cluster of hotels and transport-related services (green, upper-right), and a large metals-and-electronics cluster (blue, top-left). At the same time, there are large labor flows among industries with different colors, i.e., of different sections. However, these links across sections often connect industries that are intuitively related. For instance, high-technology manufacturing and service industries are connected in a cluster that links the *computer* and *telecommunications equipment* industries of the blue manufacturing section with the *software consulting* and *data processing* industries of the orange business services section. Similarly, we find strong links between the manufacturing industries of *printing* and *publishing* and creative services like *radio & TV* and *advertising*.

Comparing skill relatedness across labor-market segments

Does the network in Figure 2 depict a general structure or does skill relatedness differ by labor-market segment? To answer this question, we need to compare different skill-relatedness matrices to one another. We do this by first stacking all columns of a skill-relatedness matrix into one long vector. Next, we calculate the correlations among such vectors for skill-relatedness matrices of different segments.

The estimated correlations are surprisingly low, typically between 0.3 and 0.5. However, even if we compare the skill-relatedness matrix for a single segment in two consecutive years, correlations barely exceed the 0.5 mark. Although, in the long run, skill relatedness may change as technologies shift, it is implausible that it changes much on such a short horizon. This suggests that skill-relatedness matrices

15

are estimated with a substantial amount of noise. Indeed, if we assume that skill relatedness does not change from one year to the next, a year-on-year correlation of 0.5 suggests that 75% $(1 - 0.5^2)$ of the variance in the estimated skill relatedness can be attributed to random noise. As a consequence, correlations involving skill relatedness will have a strong attenuation bias, i.e., they will be biased towards zero.

To resolve this, let us consider that skill-relatedness estimates consist of two components: the (unobserved) actual skill relatedness and measurement error. Formally, we write the stacked vector of skill-relatedness estimates for labor-market segment s in year t, $\hat{r}_{st}$, as the sum of a time-invariant, real skill-relatedness vector, r_s, and a year-specific measurement error component, ϵ_{st}:

$$\hat{r}_{st} = r_s + \epsilon_{st} \tag{2}$$

If we assume that ϵ_{st} is distributed identically and independently, equation (2) implies that the bias in the correlation between two observed skill-relatedness vectors, $\hat{r}_{st}$ and $\hat{r}_{s't}$, can be reduced by averaging the skill-relatedness estimates of different years. However, because we only have nine yearly estimates, averaging $\hat{r}_{st}$ will not eliminate the attenuation bias completely. Correlations between time-averaged skill-relatedness vectors, therefore, represent, lower bounds of correlations of vectors of actual skill relatedness.

An alternative approach is to tackle the measurement error directly. This is possible, because our yearly estimates of skill relatedness, in principle, represent nine independent estimates of skill relatedness. Consequently, we can use the bias-correction method introduced by Spearman (1904):

$$Corr[r_s, r_{s'}] = \frac{Corr[\hat{r}_{st}, \hat{r}_{s't}]}{\sqrt{Corr[\hat{r}_{st}, \hat{r}_{st+1}]}\sqrt{Corr[\hat{r}_{s't}, \hat{r}_{s't+1}]}}. \tag{3}$$

16

That is, the true correlation between skill-relatedness vectors for segments s and s' can be estimated by dividing the correlation between observed skill-relatedness vectors by the square root of the correlations between two consecutive skill-relatedness measurements in each segment. Detailed derivations are provided in Appendix D.[15]

In the analyses below, we aggregate industries to the 3-digit level. This strikes a balance between the need to use relatively homogeneous industries and to maintain a large enough ratio of labor flows to industry pairs to calculate sufficiently precise skill-relatedness estimates. Using the bias-correction method of equation (3), we can now assess the extent to which skill relatedness changes over time. The bias-corrected correlation between skill relatedness in the first and last year of our data set is 0.9 or higher in all segments (Table 3).[16] This shows that, if at all, skill relatedness changes very slowly over time.

<center>TABLE 3 (CORR TIME) about here</center>

Stylized fact 3: There is no indication of rapid change in skill relatedness.

Tables 4 to 6 summarize results by labor-market segment. The first row in these tables reports the correlation for a single segment's skill relatedness in two consecutive years.[17] The higher this year-on-year own-correlation, the less noisy estimates are. The remaining rows describe to what extent the skill-relatedness measurements of a pair of segments are correlated.[18] To calculate the bottom value in

---

[15] The appendix also provides outcomes using a third bias-correction method. This approach combines the two methods described above and yields very similar results as the bias-correction method of equation (3).

[16] The measurement-error correction is in this case based on equation (4) and uses the year-on-year same-segment correlations between 1999/00 and 2000/01 and between 2006/07 and 2007/08 in the denominator.

[17] We first calculate correlations for all eight pairs of consecutive years and then take the average.

[18] Because we can only estimate the relatedness among industries with at least some inflow or outflow in each of the two labor-market segments, the number of observations varies across the cells in these tables.

these rows, we first average skill relatedness for the segments across years and then calculate the correlation of these averages. This provides a lower bound on the true correlation of two relatedness types. The upper value reports the average bias-corrected correlations using equation (3).

One potential concern is that skill-relatedness does not measure similarities in skill requirements, but simply reflects co-location patterns of industries. If this were the case, we would expect that the labor flows that take place within a region would be structurally different from those that cross into other regions. However, Table 4 shows that, at a bias-corrected correlation of 0.87, local and long-distance labor flows exhibit very similar skill-relatedness estimates. This suggests that the structure of the skill-relatedness network is not driven by industry co-location patterns.

TABLE 4 (CORR REG) about here

Stylized fact 4: Given that local and long-distance flows exhibit very similar skill-relatedness structures, skill-relatedness estimates are not mainly driven by industry co-location patterns.

Table 5 shows that, in eastern as well as in western Germany, workers with different wage levels have almost identical skill-relatedness matrices: bias-corrected correlations are all well above 0.9 and even without bias correction, we observe correlations of at least 0.8. Although differences between eastern and western Germany are slightly larger, with bias-corrected correlation estimates typically above 0.75, these differences are still relatively small.

TABLE 5 (CORR WAGES) about here

Table 6 reports correlations of skill relatedness in different occupational segments. With the exception of cleaning and security personnel, for whom bias-corrected correlations with the other occupational groupings hardly ever exceed the 0.5 mark, the different occupational groupings exhibit very similar

skill-relatedness matrices.[19] Management occupations display skill-relatedness matrices that are almost identical to those of sales people (bias-corrected estimate: 0.92), accountants (0.86), office clerks (0.91) and, to a somewhat lesser extent, IT specialists (0.83). Indeed, even without correcting for measurement error, all correlations in the first five occupational segments exceed 0.6. This stability across occupations is remarkable. Even though managers and IT specialists will carry out different tasks, they tend to switch jobs among the same industries. The fact that inter-industry flow patterns are similar for different occupations suggests that human capital has an industry-specific component that is quite independent of a worker's occupation. This contrasts with the findings of Kambourov and Manovskii (2009) Kambourov and Manovskii (2009) and Poletaev and Robinson (2008) that industry specificity is scarcely important after taking occupational specificity into account.

TABLE 6 (CORR OCC) about here

Taking the results in Tables 5 and 6 together, we arrive at the following stylized fact:

> **Stylized fact 5:** Workers with different levels and types of skills have similar skill-relatedness networks, i.e., the same industries are connected by labor flows, regardless of the skills of workers involved.

Skill relatedness and the growth of local industries

Skill relatedness can be interpreted as one specific kind of inter-industry relatedness. Inter-industry relatedness has traditionally played an important role in economic geography and the literature on

---

[19] Apparently, cleaners and security guards' job transitions are quite distinct from those of the other groups. Interestingly however, with a bias-corrected correlation of 0.79, cleaners and security guards display very similar skill-relatedness matrices. A closer inspection of their relatedness matrices suggests that even in these jobs, where industry-specific skills is arguably low, workers do not switch industries randomly. For instance, in both groups, we find strong connections among various construction industries.

geographical clusters of firms. For instance, Porter (1998) identifies local clusters of related economic activities as important sources of competitive advantages, with Silicon Valley as the archetypical example. Although originally stressing local value chains, the cluster concept has evolved to include more general linkages that "create externalities of various types" (Porter, 2003). To measure such linkages, scholars have used information on the co-location patterns of industries (Porter, 2003) and the co-occurrence of products in countries' export portfolios (Hidalgo et al., 2007). These and other relatedness measures have proven highly predictive of the growth of local industries (Neffke et al., 2011; Boschma et al., 2012; Delgado et al., 2010; Rigby, 2015; Essletzbichler, 2013; Hausmann et al., 2014). In this subsection, we investigate how skill-relatedness compares to alternative relatedness measures as a predictor of local industry growth.

In particular, we assess to what extent the presence of related industries predicts the growth and entry of local industries in German planning regions (*Raumordnungsregionen*). We define related industries on the basis of three different relatedness measures: skill relatedness, a co-location-based measure as in Porter (2003) and Hidalgo et al. (2007) and an input-output-based measure. All three measures are created using data for the period before 2003. In contrast, the analyzed growth patterns refer to the period 2003-2008.

As a measure of skill relatedness, we take the average skill-relatedness for Germany as a whole in 1999/00, 2000/01 and 2001/02. Next, we symmetrize the resulting matrix as follows:

$$SR_{ij} = \frac{\bar{R}_{ij} + \bar{R}_{ji}}{2}$$

Co-location-based relatedness is calculated as the correlation between the regional employment vectors of two industries:

$$CL_{ij} = \frac{1 + Corr(E_{ir}, E_{jr})}{2} \tag{4}$$

where E_{ir} represents the employment in industry i and region r in the year 2002. The transformation in (4) maps the correlation onto the interval [0,1]. CL_{ij} increases as the geographical distribution of employment in industries i and j becomes more similar.

For input-output relatedness, we use the German 2-digit input-output matrix of the year 2003 as provided by EUROSTAT.[20] The input-output matrix records for each pair of industries (i, j) the value of industry i's outputs purchased by industry j, V_{ij}. To establish the strength of input-output linkages between two industries, we express V_{ij} once as a percentage of all intermediate inputs into industry j and once as a percentage of all intermediate outputs of industry i. We then repeat this for the reverse flow, the value of intermediates sold by industry j to industry i. Input-output relatedness is now defined as the average of these four figures:

$$IO_{ij} = \frac{1}{4} \left(\frac{V_{ij}}{V_{i.}} + \frac{V_{ij}}{V_{.j}} + \frac{V_{ji}}{V_{j.}} + \frac{V_{ji}}{V_{.i}} \right)$$

where a dot (".") denotes summation over the omitted dimension. For all relatedness measures, we define industries to be unrelated to themselves: $SR_{ii} = CL_{ii} = IO_{ii} \equiv 0$.

We use these relatedness measures to quantify how well a local industry's related industries are represented in the region. To do so, we calculate for each industry i the weighted average employment of all other industries, $j \neq i$, in the region, using the relatedness between i and j as weights:

$$E_{irt}^{REL} = \sum_{j \neq i} \frac{REL_{ij}}{\sum_{m \neq i} REL_{im}} E_{jrt}$$

---

[20] To be precise, we use the domestic section of the product-based symmetric input-output table, ignoring international trade and sales to end-consumers.

where E_{jrt} is industry j's employment in region r and year t and REL is either SR, CL or IO.

Next, we use these variables to predict annualized growth rates of existing local industries and the entry of new industries in a region. These regressions are limited to local industries in the traded, non-resource-based, private sector.[21] Moreover, for the growth regressions, we only use industries that exist at the start of the period, estimating (cross-sectional) regression equations of the following type:

$$\log\left(\frac{E_{irt}}{E_{irt+\tau}}\right)^{\frac{1}{\tau}} = \gamma \log(E_{irt}) + \beta_{SR} \log(E_{irt}^{SR}) + \beta_{CL} \log(E_{irt}^{CL}) + \beta_{IO} \log(E_{irt}^{IO}) + \eta_i + \rho_r + \epsilon_{irt} \quad (5)$$

In equation (5), γ estimates potential mean-reversion effects. η_i and ρ_r represent industry- and region-specific effects respectively. Furthermore, t is the base year of the growth rate and τ represents the time horizon over which growth is measured.

To analyze the entry of new local industries, we run Linear Probability Models (LPMs). That is, our dependent variable is a dummy variable $ENTRY_{irt,\tau}$ that has a value of 1 if an industry i that did not yet exist in year t enters region r within the next τ years:

$$ENTRY_{irt,\tau} = \beta_{SR} \log(E_{irt}^{SR}) + \beta_{CL} \log(E_{irt}^{CL}) + \beta_{IO} \log(E_{irt}^{IO}) + \eta_i + \rho_r + \epsilon_{irt} \quad (6)$$

Like equation (5), equation (6) describes a cross-sectional analysis with fixed effects for industry and regions. However, because we only use local industries that do not yet exist in year t, there is no mean-reversion term in this equation.

Tables 7a and 7b report estimates for (5) and (6) with base years $t = 2003$ and growth and entry defined over a one-year and over a five-years period.[22] Table 7a aggregates industries to the 3-digit

---

[21] That is, we exclude industries in the 2-digit NACE classes 01-14 (agriculture, fishing and mining), 40-59 (non-traded services), and 75-99 (public sector and miscellaneous industries).

[22] Results are qualitatively similar for other base years and time windows (results available on request).

level, whereas Table 7b presents estimates for 4-digit industries. In all but one model, only the coefficients associated with skill-related employment are statistically significant. The estimated coefficients suggest that a 10% rise in skill-weighted average employment is associated with a between 1.5% and 2% increase in growth rate and between 0.5 ($0.051 * \ln(1.1)$) and 1.0 ($0.104 * \ln(1.1)$) percentage points higher entry rates. This shows that industries tend to enter and grow faster in regions with large amounts of skill-related employment. In contrast, controlling for skill-related employment, employment that is related according to co-location or input linkages typically does not show any association with growth and entry rates. In particular, skill-related employment tends to be the strongest predictor of industry growth and entry rates.

TABLE 7a AND 7b ABOUT HERE

This is not to say that co-location and input-output relatedness indicators have no merit. For one thing, collinearities among the various $\log(E_{irt}^{REL})$ terms complicate these analyses.[23] Indeed, if we regress growth and entry rates on each indicator separately all indicators exhibit some explanatory power. For another, we find stronger results for co-location-based relatedness (but not input-output based relatedness) if, instead of using the relatedness-weighted average employment, we measure related employment as all employment in related industries (where industries are considered related if they exceed a certain threshold). However, even in these regressions, skill-related employment tends to

---

[23] This is not due to high correlations among inter-industry relatedness measures. Indeed, at the 2-digit level, the correlation is 0.50 between SR_{ij} and IO_{ij}, 0.36 between SR_{ij} and CL_{ij} 0.36, and 0.40 between IO_{ij} and CL_{ij}. At the 4-digit level, these correlations drop to 0.29, 0.31 and 0.19, respectively. However, at the industry-region level, correlations among $\log(E_{irt}^{SR})$, $\log(E_{irt}^{IO})$ and $\log(E_{irt}^{CL})$ run from a maximum of 0.98 at the 2-digit level to 0.94 at the 4-digit level. Indeed, at the 2-digit level, where correlations are highest and the number of observations is lowest, multicollinearity issues become unsurmountable.

remain highly significant.[24] Overall, we therefore conclude that the predictive validity of skill-relatedness is strong vis-à-vis co-location-based and input-output relatedness:

> **Stylized fact 6:** Compared to input-output and co-location-based relatedness indices, skill relatedness is a relatively strong predictor of industries' regional growth and entry rates.

Skill relatedness and reallocation frictions

One way of thinking of the skill-relatedness patterns documented so far is that they represent constraints to labor mobility. Such constraints may hinder the efficient reallocation of workers from declining to expanding industries. This may affect a region's resilience when its economy faces adverse shocks. Indeed, just as a lack of *geographical* labor mobility makes it harder for economies to deal with asymmetric regional shocks,[25] the constrained mobility across *industries* raises similar concerns. However, the degree to which such mobility constraints will cause reallocation frictions depends on whether or not related industries experience correlated shocks. That is, only if industries that are skill related typically grow or shrink at the same time will it be hard to reallocate workers in skill-preserving ways.

To assess to what extent this is the case, we need to explore whether the differences in growth rates of two industries is associated with their skill relatedness. To do so, we first calculate skill relatedness among 3-digit industries, using only flows in the period 1999/00 to 2002/03. For the remaining years, 2004 to 2008, we calculate the absolute difference in growth rates for each pair of industries:

$$Gdif_{ij} = \left| \frac{E_{i,2008}}{E_{i,2004}} - \frac{E_{j,2008}}{E_{j,2004}} \right|$$

---

[24] Results are available upon request.

[25] See for instance, Siebert (1997) and Bentivogli and Pagano (1999) for a discussion of limited labor mobility among European regions in the context of Europe's Economic and Monetary Union as an optimal currency area.

When industries i and j grow at exactly the same rate, $Gdif_{ij}$ is zero and $Gdif_{ij}$ increases as their growth rates diverge. We find that at -0.17, the Spearman rank correlation between $\bar{R}_{ij}$ and $Gdif_{ij}$ is negative yet small. Accordingly, skill-related industries tend to have only weakly similar growth rates, which should limit potential problems for the economy as a whole to reallocate labor from shrinking to growing industries. Indeed, in Appendix E, we show that if redundant workers are reallocated to the most skill-related industry that experiences labor shortages, this reallocation can be achieved in skill-preserving ways.

In spite of this finding, however, reallocation bottlenecks may still exist in individual regions. Therefore, we repeat these analyses, but define $Gdif_{ij}$ in terms of industries' growth rates within a planning region. Figure 3 shows the histogram for the Spearman rank correlations between these regional growth differentials and skill relatedness.

<center>FIGURE 3 ABOUT HERE</center>

At the regional level, reallocation problems are even less pronounced than for Germany as a whole. For some regions rank correlations are even positive and the lowest observed rank correlation in a region is -0.126. These findings suggest the following stylized fact:

> **Stylized fact 7:** Skill-related industries typically do not exhibit highly similar growth rates. Therefore, in spite of the constrained nature of inter-industry labor flows, labor surpluses of shrinking industries can typically be absorbed by growing industries in a skill-preserving way.

5. Conclusions and future research

Our analyses of inter-industry labor flows in the German economy have yielded the following seven stylized facts. First, workers often switch jobs between industries that belong to different (highly

aggregated) sections of the industrial classification system (Stylized Fact 1). Although this fact has been interpreted as an indication that workers can change industries relatively freely and *a fortiori* that human capital cannot be particular industry specific, a closer inspection shows that industry switches are far from random. On the contrary, most labor flows take place within a narrow set of industry pairs (Stylized Fact 2): on average, 50% of job switchers move to industries that represent just 3.4% of total employment in Germany. Furthermore, after controlling for the overall flow rates of the industries involved, the underlying structure of inter-industry transitions hardly changes over time (Stylized Fact 3) and does not just reflect industries' co-location patterns (Stylized Fact 4). Moreover, job switchers with different wages and occupations follow remarkably similar flow-patterns (Stylized Fact 5).

These findings suggest that inter-industry labor flows reveal how industries are connected to one another in terms of their human-capital requirements. In line with recent literature, we therefore interpret inter-industry labor flows that exceed a well-defined random benchmark as a sign that the involved industries are skill related. Following the literature on related diversification in economic geography, we test whether our labor-flow based skill-relatedness measure predicts local industries' entry and growth rates. Indeed, the skill-relatedness index provides better growth predictions than traditional inter-industry relatedness measures based on co-location patterns or input-output relations (Stylized Fact 6). However, although the presence of skill-related industries in a region predicts growth of local industries, this does not imply that skill-related industries exhibit correlated growth patterns. Consequently, the constraints implied by the structured nature of inter-industry labor flows do not translate into major obstacles when it comes to reallocating workers from shrinking to growing industries (Stylized Fact 7).

Overall, we believe that these findings show that inter-industry labor flows are at present understudied and merit further scholarly attention. For instance, relatedness matrices such as the skill-relatedness

26

matrix in this paper have various practical and academic applications, ranging from employment and retraining programs to research in labor economics (Poletaev and Robinson, 2008; Gathmann and Schönberg, 2010), economic geography (Porter, 2003; Neffke et al., 2011; Boschma et al., 2013), development economics (Hidalgo et al., 2007) and strategic management (e.g., Teece et al., 1994; Farjoun, 1994; Bryce and Winter, 2009; Lien and Klein, 2009). Moreover, our results suggest that analyzing inter-industry flows may offer new ways to study labor markets and the flexibility of a labor force. Finally, although skill-relatedness does not change much in the relatively short period we study, if technological progress affects the skills industries require we expect to observe more drastic changes over longer time horizons. This suggests that labor-flow matrices may offer new ways to analyze technological change. We hope that the tools we present in this paper and the skill-relatedness estimates, which we make available as an appendix to this paper,[26] will allow making progress on these and other topics.

---

[26] A link to skill-relatedness estimates for all different classification systems in use in Germany between 1975 and 2014 can be found on the first author's website.

References

Becker, Gary S. 1964. Human capital. Chicago: University of Chicago Press.

Bender S, Möller J. 2010. Data from the Federal Employment Agency. In: German Data Forum & Rat für Sozial- und Wirtschaftsdaten (eds.), Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences. Vol. 2, Opladen: Budrich UniPress: 943-958.

Bentivogli C, Pagano P. 1999. Regional disparities and labour mobility: the Euro-11 versus the USA *Labour* 13(3): 737-760.

Boschma R, Minondo A, Navarro M. 2013. The emergence of new industries at the regional level in Spain. A proximity approach based on product-relatedness. *Economic Geography* 89 (1): 29-51.

Bryce DJ, Winter SG. 2009. A general inter-industry relatedness index. *Management Science* 55: 1570–1585.

Costa C, Baptista R. 2011. Agglomeration versus organizational reproduction: the molds cluster in Portugal. *Frontiers of Entrepreneurship Research* 31(4).

Dauth W 2010. The mysteries of the trade: employment effects of urban interindustry spillovers. IAB Discussion Paper 15/2010.

Delgado M, Porter, ME, Stern S. 2010. Clusters and entrepreneurship. *Journal of Economic Geography* 10(4).

Essletzbichler J 2013. Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas. *Regional Studies* DOI: 10.1080/00343404.2013.806793.

Farjoun M. 1994. Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization Science* 5 (2): 185-199.

Gartner H. 2005. The imputation of wages above the contribution limit with the German IAB employment sample. *FDZ-Methodenreport* (02/2005), Nürnberg.

Gathmann C, Schönberg U. 2010. How general is human capital? A task-based approach. *Journal of Labor Economics* 28: 1–49.

Grant RM, Spender J-C. 1996. Knowledge and the firm: overview. *Strategic Management Journal* 17: 5–9.

Grant RM. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–122.

Greenstone M, Hornbeck R, Moretti E. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(3): 536-598.

Hausmann R, Hidalgo C, Stock DP, Yildirim MA. 2014. Implied Comparative Advantage. Mimeo.

Hethey, T, Schmieder, JF. 2010. Using worker flows in the analysis of establishment turnover. Evidence from German administrative data. (FDZ Methodenreport, 06/2010), Nürnberg.

Hidalgo, CA, Klinger B, Barabási A-L Hausmann R. 2007. The product space conditions the development of nations. *Science* 317, 482-487.

Jahrbücher für Nationalökonomik and Statistik. 2016. Special Issue on 25 years of German reunification 236(2).

Kambourov G, Manovskii I. 2008. Rising occupational and industry mobility in the United States: 1968–97. *International Economic Review* 49(1): 41-79.

Lien LB Klein PG. 2009. Using competition to measure relatedness. *Journal of Management* 35(4): 1078-1107.

March JG. 1991. Exploration and exploitation in organizational learning. *Organization science* 2(1): 71-87.

Neal D. 1995. Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics* 13: 653–677.

Neffke FMH, Henning M, Boschma RA. 2011. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* 87 (3): 237-265.

Neffke FMH, Henning M. 2013. Skill-relatedness and firm diversification. *Strategic Management Journal* 34(3): 297-316

Parent D. 2000. Industry-specific capital and the wage profile: Evidence from the national longitudinal survey of youth and the panel of income dynamics. *Journal of Labor Economics* 18: 306–323.

Parrado E, Caner A, Wolff EN. 2007. Occupational and industrial mobility in the United States. *Labour Economics* 14(3): 435–455.

Poletaev M, Robinson C. (2008): Human Capital Specificity: Evidence from the dictionary of occupational titles and displaced worker surveys. *Journal of Labour Economics* 26 (3):387-420.

Porter ME. 1987. From competitive advantage to corporate strategy. *Harvard Business Review* 65(3):43-59.

Porter ME. 1998. Clusters and the new economics of competition. Harvard Business Review 76(November-December): 77-90.

Porter ME. 2003. The economic performance of regions. *Regional Studies* 37(6-7): 549-578.

Kaplan DS, Lederman D, Robertson R. 2011. Employment and Wage Responses to Trade Shocks: Evidence from Mexico during the 2008-09 US Recession, mimeo.

Rigby DL. 2015. Technological Relatedness and Knowledge Space: Entry and Exit of U.S. Cities from Patent Classes. Regional Studies 49(11): 1922-1937.Saxenian A. 2006. The new Argonauts: Regional advantage in the global economy. Cambridge, MA: Harvard University Press.

Siebert H. 1997. Labor market rigidities: At the root of unemployment in Europe. *Journal of Economic Perspectives* 11(3):37-54.

Simon H. 1991. Bounded rationality and organizational learning. *Organization Science* 2: 125-134.

Spearman C. 1910. Correlation calculated from faulty data. *British Journal of Psychology* 3(3): 271-295.

Sullivan P. 2010. Empirical evidence on occupation and industry specific human capital. *Labour Economics* 17(3): 567–580.

Teece DJ, Rumelt R, Dosi G, Winter S. (1994): Understanding corporate coherence: Theory and Evidence. *Journal of Economic Behaviour and Organization* 23: 1-30.

Timmermans B, Boschma R. 2014. The effect of intra- and inter-regional labour mobility on plant performance in Denmark: the significance of related labour inflows. *Journal of Economic Geography 14(2):289-311.*

Winkelmann U. 2010 „Manche pendeln weit" – Berufspendler im Bundesländervergleich *Statistisches Monatsheft Baden-Württemberg* 2010(4): 40-43.

Appendix A: Classification systems

The German industry classification system has seen three major overhauls since 1975. From 1975 to 1998 the Statistical Classification of Economic Activities in Germany 1973 (WZ 1973) was used. In 1999 the WZ 1993 was implemented and used until 2003, followed by WZ 2003 (used from 2003-2008) and the WZ 2008 in 2008. The WZ 1993, WZ 2003 and WZ 2008 were harmonized at the 4-digit level with the European NACE (Nomenclature générale des Activités économiques dans les Communautés Européennes) 1.0, NACE Revision 1.1 and NACE Revision 2.0 classifications. The classification systems in the period of our study, WZ 1993 and WZ 2003, consist of six levels. Industries are first grouped into sets coded by two capital letters, the first of which denotes the section to which the industry belongs and the second divides these sections into sub-sections. The remaining four levels are indicated by the first 2, 3, 4 and 5 digits of a numeric code.

With the introduction of the WZ 2003 a minor reclassification took place. Therefore we harmonize the WZ 1993 and WZ 2003, taking the WZ 2003 as a starting point. For plants that exist before 2003, we apply the WZ 2003 industry codes to earlier years, provided that the establishment's WZ 1993 code remains the same. Next, we construct a concordance between WZ 1993 and WZ 2003 to fill in the gaps that occur when plants exit or change their WZ 1993 codes before 2003. This concordance is based on information from the year 2003 in which establishments' industries are recorded in both classification systems. For most 5-digit WZ 1993 industry-codes, we find all but unique translations to the WZ 2003 system. However, in some cases single WZ 1993 industries were split or multiple WZ 1993 were merged, we merge 60 out of 1043 WZ 2003 classes into 29 combined industries. The exact procedures are coded in the STATA do-files accompanying this paper.

Appendix B: Identifying spurious labor flows

Hethey and Schmieder (2010) argue that the reliability of the establishment identifiers in the HES data can be improved by analyzing labor flows among establishments. For instance, if all workers in a new establishment come from the same other establishment, this establishment is likely to be a spinoff or an existing establishment with a changed code. To separate real labor flows from the spurious flows that arise from recoding establishments, we analyze inter-establishment labor flows. Whether a flow should be considered spurious depends on the size of the establishments of the flow's origin and destination.

For establishments with fewer than five employees, a flow is considered spurious if all employees either came from or went to the same establishment (the exception are establishments with only one employee). For plants with five or more employees, flows are considered spurious if they represent at least 30% of the employment in the establishment where the flow originates or ends. Furthermore, flows of 100 employees or more are considered spurious regardless of establishment sizes. We illustrate these choices in Figures B1 and B2, which depict the yearly average number of establishment-to-establishment transitions that are excluded at different thresholds.

Figure B1 shows how many individuals in establishments with at least 5 employees are dropped. A threshold of 30% (the lower bound used by Hethey and Schmieder, 2010) drops about 27.5% of the about 1.8 million switchers. This percentage does not change much until the cutoff exceeds 60%. Figure B2 plots the number of individuals dropped for smaller establishments. Because small establishments represent only a modest number of job switchers, not many flows are lost regardless of the chosen threshold. The imposition of a 100% threshold in this paper implies that about 32,200 establishment switchers are disregarded per year. The final restriction we impose is that inter-establishment flows should never consists of blocks larger than 100 individuals. This condition reduces the sample by another 11,100 individuals per year.

Figure B1: Suspect flows (large establishments)
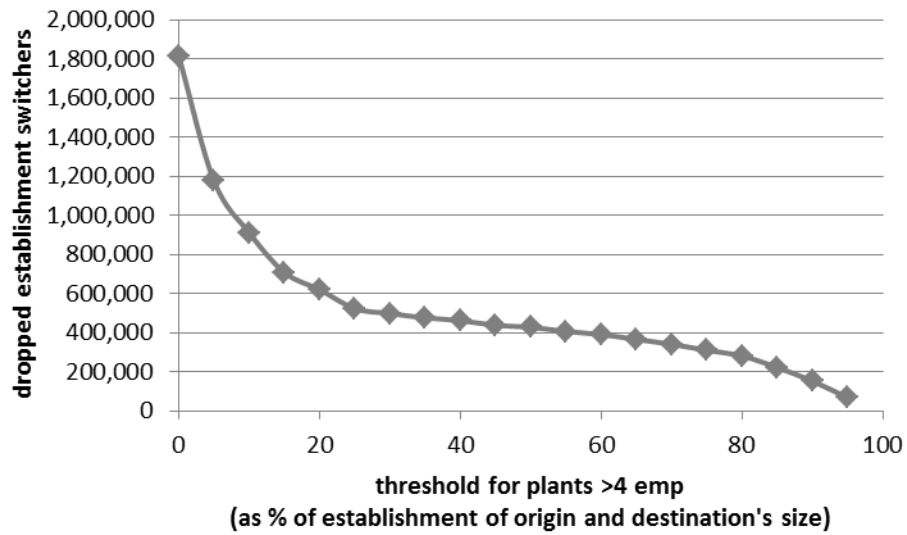
Number of job switchers that are excluded in establishments with 5 or more employees at different thresholds.

Figure B2: Suspect flows (small establishments)
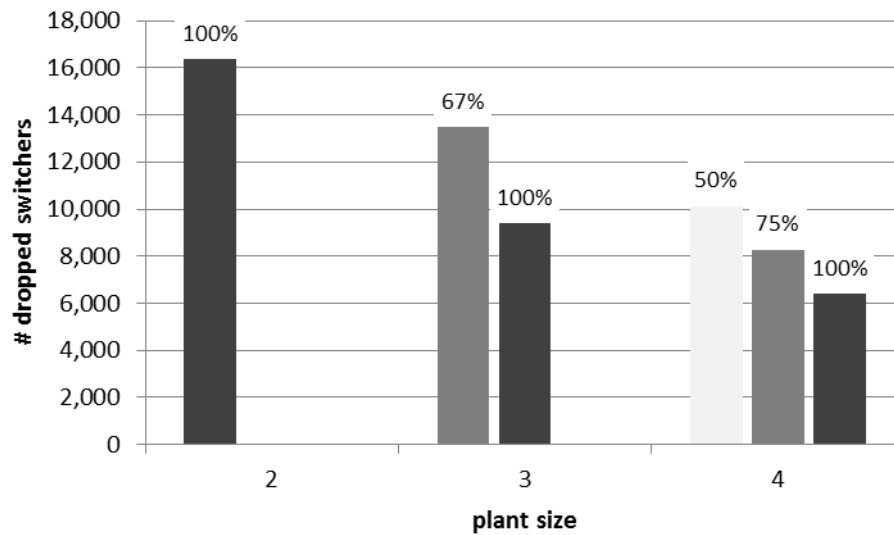
Bars represent the number of switchers that are excluded at the threshold depicted above each bar.

Appendix C: Occupational groupings

Table C: Number of employees by selected occupations and occupational grouping

Occupations	average # emp. (1999-2007)
Management Occupations	**441,926**
751: Entrepreneurs, managing directors, divisional managers	341,450
762: Executive administrative professionals	100,476
Sales Occupations	**1,448,132**
681: Wholesale and retail trade buyers, buyers	367,342
682: Salespersons	772,100
687: Commercial agents, travelers	175,604
701: Forwarding business dealers	73,839
703: Publicity occupations	59,247
Accountants	**301,287**
753: Chartered accountants, tax advisers	114,940
771: Cost accountants, valuers	37,693
772: Accountants	148,654
Office clerks	**2,653,181**
781: Office specialists	2,653,181
IT specialists	**399,136**
774: Data processing specialists	399,136
Cleaners	**118,921**
793: Doormen, caretakers	145,238
933: Household cleaners	193,222
934: Glass, buildings cleaners	38,508
Security personnel	**231,730**
791: Factory guards, detectives	17,977
792: Watchmen, custodians	100,944

Occupations are classified in 3-digit codes by the German Classifications of Occupations 1973.

Appendix D: Derivation attenuation bias correction

In this appendix, we derive the bias-correction equation of equation (3). First, $r_{t,s}$ as the stacked the $\bar{R}$-matrix of a labor-market segment s in year t into a vector, omitting diagonal elements:

$$r_{t,s} = \begin{pmatrix} \bar{R}_{1,2}^{t,s} \\ \vdots \\ \bar{R}_{1,n}^{t,s} \\ \vdots \\ \bar{R}_{n,1}^{t,s} \\ \vdots \\ \bar{R}_{n,(n-1)}^{t,s} \end{pmatrix}$$

In this vector, $\bar{R}_{ij}^{t,s}$ represents the skill relatedness between industries i and j in year t and segment s. As described in equation (2), relatedness estimates can be thought of as consisting of two components: a structural, yet unobserved component and an orthogonal random component.

$$\hat{r}_{st} = r_s + \varepsilon_{st}$$

where $\hat{r}_{st}$ represents the observed vector of relatedness estimates, r_s the true relatedness, and ε_{st} a measurement error. Note that r_s does not depend on t. That is, we assume that the true underlying relatedness is unchanging.

For notational clarity, we denote the estimated skill-relatedness vector in segment s in year t, $\hat{r}_{st}$, and its true vector, r_s, by $\hat{x}_t$ and x. Similarly, $\hat{y}_t$ and y_t correspond to the estimated and real skill-relatedness vectors in segment s'. The correlation between the skill-relatedness estimates of the two segments can now be written as:

$$Corr[\hat{x}_t, \hat{y}_t] = \frac{Cov[x+\varepsilon_{xt}, y+\varepsilon_{yt}]}{\sqrt{Var[x+\varepsilon_{xt}]Var[y+\varepsilon_{yt}]}} \tag{D1}$$

Or,

$$Corr[\hat{x}_t, \hat{y}_t] = \frac{Cov[x,y]+Cov[x,\varepsilon_{yt}]+Cov[\varepsilon_{xt},y]+Cov[\varepsilon_{xt},\varepsilon_{yt}]}{\sqrt{(Var[x]+2Cov[x,\varepsilon_{xt}]+Var[\varepsilon_{xt}])(Var[y]+2Cov[y,\varepsilon_{yt}]+Var[\varepsilon_{yt}])}}$$

Let us now assume that the measurement errors are uncorrelated with the true skill-relatedness values:

Assumption 1a: $Corr(x, \varepsilon_{xt}) = Corr(y, \varepsilon_{yt}) = 0$

Assumption 1b: $Corr(x, \varepsilon_{yt}) = Corr(y, \varepsilon_{xt}) = 0$

Because skills are latent constructs, Assumption 1a can be interpreted as a definition: whatever it is that we will refer to as skill relatedness, its estimated value can be decomposed into an invariant, structural term and into an error term. Let us further assume that the error terms for both relatedness types are uncorrelated as well:

Assumption 2: $Corr(\varepsilon_{xt}, \varepsilon_{yt}) = 0$

Using the assumptions, we can rewrite (D1) as:

$$Corr[\hat{x}_t, \hat{y}_t] = \frac{Cov[x,y]}{\sqrt{(Var[x]+Var[\varepsilon_{xt}])(Var[y]+Var[\varepsilon_{yt}])}}$$

or, multiplying by $\frac{\sqrt{Var[x]Var[y]}}{\sqrt{Var[x]Var[y]}}$:

$$Corr[\hat{x}_t, \hat{y}_t] = \frac{Cov[x,y]}{\sqrt{Var[x]Var[y]}} \frac{\sqrt{Var[x]}}{\sqrt{Var[x]+Var[\varepsilon_{xt}]}} \frac{\sqrt{Var[y]}}{\sqrt{Var[y]+Var[\varepsilon_{yt}]}}$$

which is equation (9) in the main text. Rearranging terms and using the fact that $\rho_{xy} = Corr(x,y) = \frac{Cov[x,y]}{\sqrt{Var[x]Var[y]}}$ yields the following expression for the real correlation across segments:

$$\rho_{xy} = \rho_{\hat{x}_t\hat{y}_t} \frac{\sqrt{Var[x]+Var[\varepsilon_{xt}]}}{\sqrt{Var[x]}} \frac{\sqrt{Var[y]+Var[\varepsilon_{yt}]}}{\sqrt{Var[y]}} = \rho_{\hat{x}_t\hat{y}_t} \sqrt{1 + \frac{Var[\varepsilon_{xt}]}{Var[x]}} \sqrt{1 + \frac{Var[\varepsilon_{yt}]}{Var[y]}} \qquad (D2)$$

(D2) shows that the downward bias in the measured correlation depends on the relative sizes of the error variances compared to the variance of the relatedness types. Therefore, we need an estimate of the relative size of the measurement errors. To arrive at such an estimate, we will assume that the error terms are uncorrelated over time.

Assumption 3a: $\forall_{t,t'}: t \neq t' \rightarrow Corr(\varepsilon_{xt}, \varepsilon_{xt'})=0$

Assumption 3b: $\forall_{t,t'}: t \neq t' \rightarrow Corr(\varepsilon_{yt}, \varepsilon_{yt'})=0$

The correlation between two measurements of the same skill-relatedness type can be written as:

$$Corr[\hat{x}_t, \hat{x}_{t'}] = \frac{Cov[x+\varepsilon_{xt}, x+\varepsilon_{xt'}]}{\sqrt{Var[x+\varepsilon_{xt}]Var[x+\varepsilon_{xt'}]}}$$

$$Corr[\hat{x}_t, \hat{x}_{t'}] = \frac{Cov[x,x]+Cov[x,\varepsilon_{xt'}]+Cov[\varepsilon_{xt},x]+Cov[\varepsilon_{xt},\varepsilon_{xt'}]}{\sqrt{(Var[x]+2Cov[x,\varepsilon_{xt}]+Var[\varepsilon_{xt}])(Var[x]+2Cov[x,\varepsilon_{xt'}]+Var[\varepsilon_{xt'}])}}$$

Using Assumptions 3a and 3b, we arrive at:

$$Corr[\hat{x}_t, \hat{x}_{t'}] = \frac{Var[x]}{\sqrt{(Var[x]+Var[\varepsilon_{xt}])(Var[x]+Var[\varepsilon_{xt'}])}}$$

We will also assume that the measurement error has about the same variance in different years:

Assumption 4a: $\forall_{t,t'}: Var(\varepsilon_{xt}) = Var(\varepsilon_{xt'}) = Var(\varepsilon_x)$

Assumption 4b: $\forall_{t,t'}: Var(\varepsilon_{yt}) = Var(\varepsilon_{yt'}) = Var(\varepsilon_y)$

$$Corr[\hat{x}_t, \hat{x}_{t'}] = \frac{Var[x]}{Var[x]+Var[\varepsilon_x]} \tag{D3}$$

Similar derivations for y yield:

$$Corr[\hat{y}_t, \hat{y}_{t'}] = \frac{Var[y]}{Var[y]+Var[\varepsilon_y]} \tag{D4}$$

Denoting correlation by ρ, we arrive at (3) by substituting (D3) and (D4) into (D2):

$$\rho_{xy} = \rho_{\hat{x}_t \hat{y}_t} \frac{1}{\sqrt{\rho_{\hat{x}_t, \hat{x}_{t\prime}}}} \frac{1}{\sqrt{\rho_{\hat{y}_t, \hat{y}_{t\prime}}}} = \frac{\rho_{\hat{x}_t \hat{y}_t}}{\sqrt{\rho_{\hat{x}_t, \hat{x}_{t\prime}}} \sqrt{\rho_{\hat{y}_t, \hat{y}_{t\prime}}}}$$

The impact of measurement error can also be reduced by first averaging relatedness estimates over several years and then calculating correlations among these averaged values. Because we only have a finite number of years of observations, some attenuation bias will remain. This method, therefore, provides a lower bound for the true correlations. A third approach combines both corrections. It first averages out part of the measurement error and then corrects for any remaining bias.

Let $\bar{x} = \frac{1}{T}\sum_t \hat{x}_t$ and $\bar{y} = \frac{1}{T}\sum_t \hat{y}_t$ be the average across T years of observations of the skill-relatedness types x and y. The correlation between these two averages can be written as:

$$Corr[\bar{x}, \bar{y}] = \frac{Cov[\frac{1}{T}\sum(x+\varepsilon_{xt}), \frac{1}{T}\sum(y+\varepsilon_{yt\prime})]}{\sqrt{Var[\frac{1}{T}\sum(x+\varepsilon_{xt})]Var[\frac{1}{T}\sum(y+\varepsilon_{yt})]}}$$

$$Corr[\bar{x}, \bar{y}] = \frac{Cov[\sum(x+\varepsilon_{xt}), \sum(y+\varepsilon_{yt\prime})]}{\sqrt{Var[\sum(x+\varepsilon_{xt})]Var[\sum(y+\varepsilon_{yt})]}} \tag{D5}$$

Assuming constant variances for the measurement errors (Assumptions 4a and 4b), the numerator of (D5) can be written as:

$$Cov[\sum(x+\varepsilon_{xt}), \sum(y+\varepsilon_{yt\prime})] = T^2 Cov[x,y] + T\sum Cov[x, \varepsilon_{yt}] + T\sum Cov[\varepsilon_{xt}, y] + \sum\sum Cov[\varepsilon_{xt}, \varepsilon_{yt\prime}]$$

By assumptions (1a), (1b), (2), (3a), (3b), (4a) and (4b), this simplifies to:

$$Cov[\sum(x+\varepsilon_{xt}), \sum(y+\varepsilon_{yt\prime})] = T^2 Cov[x,y] \tag{D6}$$

Expanding the first term of the denominator of (D5), we derive the following:

$$Var[\sum(x+\varepsilon_{xt})] = T^2 Var[x] + TVar[\varepsilon_{xt}] + \sum_{t\neq t\prime}\sum_t Cov[\varepsilon_{xt}, \varepsilon_{xt\prime}] + \sum\sum Cov[x, \varepsilon_{xt}]$$

which simplifies under the abovementioned assumption to:

$$Var[\sum(x + \varepsilon_{xt})] = T^2 Var[x] + T Var[\varepsilon_x] \tag{D7}$$

Due to similar considerations for y, the second term in the denominator of (D5) is:

$$Var[\sum(x + \varepsilon_{yt})] = T^2 Var[y] + T Var[\varepsilon_y] \tag{D8}$$

Substituting (D6), (D7)and (D8) into (D5)(D4), we get:

$$Corr[\bar{x}, \bar{y}] = \frac{T^2 Cov[x,y]}{\sqrt{T^2 Var[x] + T Var[\varepsilon_x]}\sqrt{T^2 Var[y] + T Var[\varepsilon_y]}} = \frac{Cov[x,y]}{\sqrt{Var[x] + \frac{1}{T}Var[\varepsilon_x]}\sqrt{Var[y] + \frac{1}{T}Var[\varepsilon_y]}} \tag{D9}$$

Rearranging (D4) yields the following expression for $Var[\varepsilon_x]$:

$$Var[\varepsilon_x] = \frac{Var[x]}{Corr[\hat{x}_t, \hat{x}_{t\prime}]} - Var[x] = \frac{1 - Corr[\hat{x}_t, \hat{x}_{t\prime}]}{Corr[\hat{x}_t, \hat{x}_{t\prime}]} Var[x] \tag{D10}$$

Substituting (D10) and its counterpart for y into (D9) yields:

$$Corr[\bar{x}, \bar{y}] = \frac{Cov[x,y]}{\sqrt{Var[x] + \frac{1}{T}\frac{1 - Corr[\hat{x}_t, \hat{x}_{t\prime}]}{Corr[\hat{x}_t, \hat{x}_{t\prime}]} Var[x]}\sqrt{Var[y] + \frac{1}{T}\frac{1 - Corr[\hat{y}_t, \hat{y}_{t\prime}]}{Corr[\hat{y}_t, \hat{y}_{t\prime}]} Var[y]}}$$

Now we can rearrange the terms to arrive at:

$$Corr[\bar{x}, \bar{y}] = \frac{Cov[x,y]}{\sqrt{Var[x]}\sqrt{Var[y]}} \frac{1}{\sqrt{1 + \frac{1}{T}\frac{1 - Corr[\hat{x}_t, \hat{x}_{t\prime}]}{Corr[\hat{x}_t, \hat{x}_{t\prime}]}}} \frac{1}{\sqrt{1 + \frac{1}{T}\frac{1 - Corr[\hat{y}_t, \hat{y}_{t\prime}]}{Corr[\hat{y}_t, \hat{y}_{t\prime}]}}}$$

which can be rewritten as:

$$\rho_{xy} = \rho_{\bar{x}\bar{y}}\sqrt{1 + \frac{1}{T}\frac{1 - \rho_{\hat{x}_t \hat{x}_{t\prime}}}{\rho_{\hat{x}_t \hat{x}_{t\prime}}}}\sqrt{1 + \frac{1}{T}\frac{1 - \rho_{\hat{y}_t \hat{y}_{t\prime}}}{\rho_{\hat{y}_t \hat{y}_{t\prime}}}} \tag{D11}$$

Equation (D11), which was first proposed by Spearman (1910), and equation (3) in the main text will give the same results (in expected terms) as long as none of the assumptions are violated. This provides a convenient check for our empirical results. To test this, Tables D1 to D3 replicate Tables 4 to 6. Instead

of reporting the correlation between time-averaged skill-relatedness matrices in the bottom rows of the cells, we now adjust these correlations using equation (D11). The top row in each cell is still based on the bias-correction of equation (3). In almost all cases, the bias-corrections in (D11) and (3) give almost identical results, adding credence to the bias-corrected estimates we report in the main text.

Table D1: Comparison bias-correction methods: local versus long-distance flows

	local	long-distance
own correlation	0.56	0.51
local	n.a.	
	n.a.	
long-distance	0.87	n.a.
	0.88	n.a.

Idem Table 4. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).

Table D2: Comparison bias-correction methods: Wages

	high East	low East	high West	low West
own correlation	0.45	0.43	0.56	0.51
high East	n.a.			
	n.a.			
low East	0.94	n.a.		
	0.92	n.a.		
high West	0.79	0.74	n.a.	
	0.80	0.77	n.a.	
low West	0.75	0.79	0.93	n.a.
	0.76	0.82	0.92	n.a.

Idem Table 5. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).

Table D3: Comparison bias-correction methods: Occupations

	managers	sales	accountants	office clerks	IT	security	cleaners
own correlation	0.42	0.43	0.29	0.44	0.33	0.34	0.27
managers	n.a.						
	n.a.						
sales	0.92	n.a.					
	0.89	n.a.					
accountants	0.86	0.80	n.a.				
	0.83	0.78	n.a.				
office clerks	0.91	0.89	0.78	n.a.			
	0.89	0.87	0.75	n.a.			
IT	0.83	0.78	0.89	0.76	n.a.		
	0.83	0.77	0.86	0.74	n.a.		
security	0.44	0.38	0.53	0.39	0.44	n.a.	
	0.46	0.38	0.53	0.41	0.43	n.a.	
cleaners	0.50	0.41	0.60	0.47	0.57	0.79	n.a.
	0.49	0.43	0.59	0.46	0.50	0.79	n.a.

Idem Table 6. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).

Appendix E: Reallocation frictions

In this appendix we explore to what extent workers can be reallocated from shrinking to growing industries in skill-preserving ways. We define the reallocation potential from a shrinking industry i to a growing industry j as the minimum of the number of workers made redundant in the shrinking industry and the employment expansion in the growing industry. In particular, if D_i is the (potentially negative) net labor demand that industry i accumulates between 2004 and 2008, i.e., its employment growth between 2004 and 2008, the reallocation potential from industry i to industry j is:

$$POT_{ij} = \min(-D_i, D_j) \; if \; D_i < 0 \; and \; D_j > 0$$
$$= 0 \qquad \qquad otherwise$$

The reallocation problem the economy needs to resolve is to create *net* labor flows such that the labor demand of growing industries equals the labor surplus of shrinking industries. Ideally, this reallocation would shift workers between highly skill-related industries.[27] We now explore at which degree of skill relatedness workers could have been reallocated, had labor shortages been resolved "optimally",[28] that is, by reallocating labor in decreasing order of skill relatedness. To disregard overall growth, we first scale shortages by a common factor such that overall shortages and surpluses sum to zero.[29] Next, we move down the list of industry pairs in descending order of skill relatedness, shifting workers from

---

[27] In this context, the correlation between reallocation potential and skill relatedness is more important than the correlation between growth rates reported in the main text. The reason is that reallocation potentials reflect the *size* of the labor surplus that needs to be reallocated, whereas the difference in growth rates does not take industries' absolute sizes into account. Positive correlations imply that reallocation potentials are larger for more skill-related industries. Indeed, none of the German spatial planning regions displays significantly negative correlations between industries' skill relatedness and their local reallocation potential.

[28] In principle, trading off lower skill relatedness in some industry pairs against higher skill relatedness elsewhere might lead to even higher average levels of skill relatedness. However, to make such trade-offs, skill relatedness needs to have a clear ratio-scale interpretation, which is not guaranteed.

[29] Essentially, this assumes that all growing industries meet their increased labor demand to the same extent with new workers and that the downsizing in all shrinking industries relies to the same extent on workers exiting the labor market.

shrinking to growing industries until all labor surpluses are exhausted. The average $\bar{R}_{ij}$ of these artificial reallocation flows is 0.415, substantially higher than the average $\bar{R}_{ij}$ for actual job switches (0.081) in the same period, suggesting that industries' changing labor demands could have been easily met in skill-preserving ways.

To identify potential frictions at the level of local labor markets, we repeat the reallocation thought-experiment for Germany's 96 spatial planning regions (i.e., *Raumordnungsregionen*). Figure E1 shows that, in all regions, the average $\bar{R}_{ij}$ of artificial reallocation flows exceeds the average $\bar{R}_{ij}$ for actual job switches by far, reiterating the conclusion in Stylized Fact 7 that, also at the regional level, the constraints to inter-industry mobility do not need to negatively impact the efficient reallocation of labor resources.
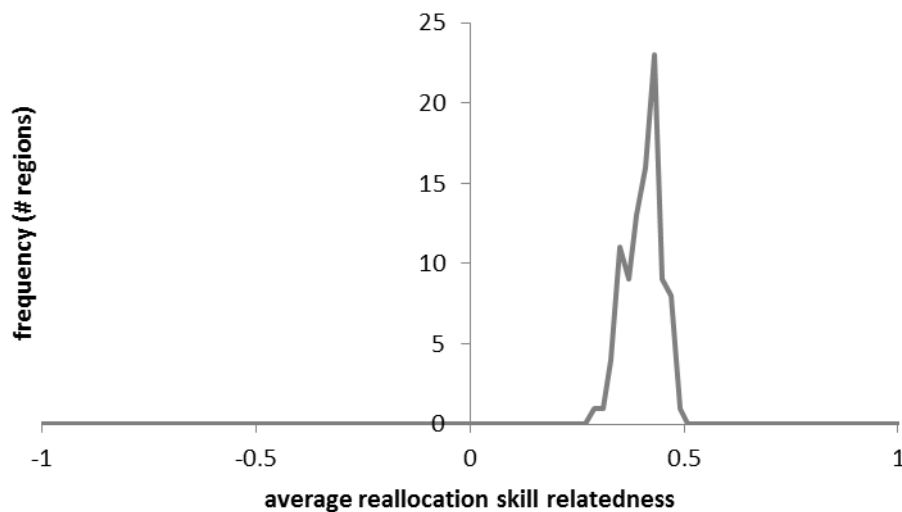

Figure E1: Regional labor reallocation: average skill relatedness by region

The graph shows the histogram of the average skill relatedness at which workers can be reallocated from shrinking to growing industries in regions.

Tables

Table 1: Cross-industry labor flows by labor-market segment

WORKERS (THOUSANDS)

LABOR-MARKET SEGMENT	WAGES			GERMANY		GEOGRAPHY		OCCUPATIONS						
	all	high	low	East	West	local	long-distance	managers	sales	accountants	office clerks	IT	cleaners	security
employment	19,897.1	9,947.6	9,914.8	3,890.2	15,958.8	n.a.	n.a.	435.9	1,440.2	293.1	2,600.8	390.5	220.2	254.2
job switchers	1,206.7	554.3	652.4	194.2	940.9	948.2	258.5	32.0	116.5	18.9	148.5	28.7	11.1	14.7
no industry switch	321.1	165.2	155.9	56.2	246.1	255.3	65.8	7.1	32.4	5.3	32.8	4.5	3.3	4.9
industry switch	885.7	389.1	496.5	138.0	694.8	692.9	192.8	24.9	84.1	13.7	115.7	24.2	7.8	9.8
different sector	519.8	215.9	303.8	82.1	408.0	409.3	110.4	14.5	45.2	8.0	75.4	14.1	6.1	7.2
same sector	365.9	173.2	192.7	56.0	286.8	283.6	82.3	10.3	38.9	5.7	40.3	10.0	1.7	2.6
same sub-sector	301.6	145.4	156.2	50.1	231.1	229.8	71.7	8.6	36.2	5.2	34.5	9.4	1.6	2.5
same 2-digit industry	225.9	109.8	116.0	40.4	170.4	175.3	50.6	6.0	25.6	4.0	23.4	4.6	1.3	1.9
same 3-digit industry	117.3	58.8	58.5	20.5	88.6	90.4	26.9	2.8	11.4	3.0	10.8	2.1	0.3	0.6
same 4-digit industry	62.3	31.7	30.6	10.4	47.6	47.6	14.8	1.4	6.0	1.5	5.1	1.7	0.2	0.3
PERCENTAGES														
no industry switch	26.6%	29.8%	23.9%	28.9%	26.2%	26.9%	25.4%	22.2%	27.8%	27.9%	22.1%	15.7%	29.6%	33.3%
industry switch	73.4%	70.2%	76.1%	71.1%	73.8%	73.1%	74.6%	77.8%	72.2%	72.1%	77.9%	84.3%	70.4%	66.7%
different sector	58.7%	55.5%	61.2%	59.5%	58.7%	59.1%	57.3%	58.4%	53.8%	58.4%	65.2%	58.5%	78.8%	73.9%
same sector	41.3%	44.5%	38.8%	40.5%	41.3%	40.9%	42.7%	41.6%	46.2%	41.6%	34.8%	41.5%	21.2%	26.1%
same sub-sector	34.1%	37.4%	31.4%	36.3%	33.3%	33.2%	37.2%	34.7%	43.1%	38.0%	29.8%	38.8%	20.2%	25.7%
same 2-digit industry	25.5%	28.2%	23.4%	29.3%	24.5%	25.3%	26.3%	23.9%	30.4%	29.3%	20.3%	19.0%	16.2%	19.1%
same 3-digit industry	13.2%	15.1%	11.8%	14.9%	12.7%	13.0%	14.0%	11.3%	13.5%	21.7%	9.3%	8.7%	4.4%	5.6%
same 4-digit industry	7.0%	8.1%	6.2%	7.5%	6.8%	6.9%	7.7%	5.7%	7.1%	10.9%	4.4%	6.8%	3.0%	3.3%

The table contains information about absolute and relative employment levels, jobs switching and cross-industry flows in each labor-market segment averaged over 1999-2007. In the section WAGES, column <all> refers to all workers, column <high> refers to workers with wages above their industry's median, column <low> to workers below this median. Long-distance flows are labor flows between establishments that are at least 100 km apart. The occupations are as defined in Table C1.

Table 2: $FLEX(0.50)$ estimates by labor-market segment

segment	$FLEX(50)$ (inc. intra-ind flows)	$FLEX(50)$ (exc. intra-ind flows)
Germany		
all	3.4% (33.2%)	5.9% (33.0%)
high wage	2.6% (22.8%)	4.2% (22.6%)
low wage	3.4% (31.6%)	5.9% (31.4%)
East / West Germany		
East	4.0% (29.1%)	6.4% (28.9%)
West	2.2% (28.0%)	3.9% (27.7%)
Geography of flows		
local	3.4% (31.8%)	6.0% (31.6%)
long-distance	2.3% (22.3%)	3.4% (22.1%)
Occupations		
managers	2.3% (8.5%)	2.9% (8.4%)
sales	3.2% (25.2%)	4.6% (24.5%)
accountants	3.2% (25.2%)	4.6% (24.5%)
office clerks	3.8% (18.8%)	15.0% (22.8%)
IT	16.8% (26.0%)	5.6% (18.4%)
cleaners	16.5% (30.3%)	6.2% (18.4%)
security	11.9% (22.0%)	8.9% (16.7%)

$FLEX(0.50)$ equals the employment weighted average employment share (across industries of origin) of the 5-digit destination industries with the highest labor-flow-to-employment ratios that taken together absorb 50% of all labor flows originating from a given 5-digit industry. All values represent averages across the period 1999-2007. The number provided in parenthesis represents the $FLEX(0.50)$ that was simulated under the null-model.

Table 3: Changes in skill relatedness between 1999 and 2008

segment	stability
Germany	
all wages	0.93
high wages	0.93
low wages	0.93
East / West Germany	
East	0.90
West	0.93
Geography of flows	
local	0.93
long-distance	0.95
Occupations	
managers	0.94
sales	0.97
accountants	0.92
office clerks	0.95
IT	0.93
cleaners	0.91
security	0.93

Stability is the attenuation-bias-corrected correlation between skill relatedness in year 1999 and 2007 of a given labor-market segment. Measurement error variances are assessed by the correlations between skill-relatedness estimates for 1999/00 and 2000/01 and for 2006/07 and 2007/08 of the corresponding labor-market segment.

Table 4: Correlations of skill-relatedness estimates of local versus long-distance flows

	local	long-distance
own correlation	0.56	0.51
local	n.a.	
	1.00	
long-distance	0.87	n.a.
	0.81	1.00

The table presents correlations between the skill-relatedness ($\bar{R}_{ij}$) matrices for the combination of labor-market segments given in the rows and columns. The first row of the table provides the average correlation between skill-relatedness matrices for two consecutive years of the labor-market segment in the columns. In the remaining cells, the top row represents the bias-corrected correlation using equation (3) and the bottom row the correlation between relatedness matrices averaged over all available years. Industries have been aggregated to the 3-digit level. Local flows are flows taking place over a road-distance of below 100 km, flows beyond this distance are labelled long-distance.

Table 5: Correlations of skill-relatedness estimates by wage levels and region

	high East	low East	high West	low West
own correlation	0.45	0.43	0.56	0.51
high East	n.a. 1.00			
low East	0.94 0.80	n.a. 1.00		
high West	0.79 0.72	0.74 0.69	n.a. 1.00	
low West	0.75 0.67	0.79 0.73	0.93 0.84	n.a. 1.00

Idem Table 4. High (low) East (West) represents high-wage (low-wage) workers in eastern (western) Germany.

Table 6: Correlations of skill-relatedness estimates for different occupations

	managers	sales	accountants	office clerks	IT	security	cleaners
own correlation	0.42	0.43	0.29	0.44	0.33	0.34	0.27
managers	n.a. 1.00						
sales	0.92 0.77	n.a. 1.00					
accountants	0.86 0.69	0.80 0.64	n.a. 1.00				
office clerks	0.91 0.78	0.89 0.76	0.78 0.62	n.a. 1.00			
IT	0.83 0.69	0.78 0.63	0.89 0.68	0.76 0.60	n.a. 1.00		
security	0.44 0.39	0.38 0.33	0.53 0.43	0.39 0.35	0.44 0.37	n.a. 1.00	
cleaners	0.50 0.39	0.41 0.35	0.60 0.45	0.47 0.36	0.57 0.38	0.79 0.62	n.a. 1.00

Idem Table 4. Row and column labels refer to broad occupational groupings.

Table 7a: Local industry growth and entry regressions (3-digit industries)

	(1) growth 2003-2004	(2) growth 2003-2008	(3) entry 2003-2004	(4) entry 2003-2008
log(Eirt)	-0.035***	-0.030***		
	(0.0036)	(0.0015)		
log(Eirt_SR)	0.171***	0.158***	0.074*	-0.061
	(0.0416)	(0.0169)	(0.0399)	(0.0705)
log(Eirt_CL)	0.126	0.082	-0.152	1.448*
	(0.1911)	(0.0854)	(0.5325)	(0.8588)
log(Eirt_IO)	0.004	-0.016	-0.003	-0.077
	(0.0234)	(0.0100)	(0.0264)	(0.0538)
R-squared	0.054	0.175	0.226	0.253
N	12,408	12,168	2,046	2,046

***: $p<0.01$, **: $p<0.05$, *: $p<0.10$, robust standard errors in parentheses. All models include region and industry fixed effects. Regional units are defined as Germany's planning regions and industries are aggregated at the 3-digit level, excluding industries in agriculture, fishing and mining (NACE 01-14), non-traded industries (NACE 40-59) and public sector industries (NACE 75-99). Columns (1) and (2) report OLS regressions of the logarithm of annualized growth rates over a one- and over a five-year period. Columns (3) and (4) report the outcomes of OLS regressions with as a dependent variable a dummy variable that evaluates whether or not a new local industry enters a region within a one- or a five-year period.

Table 7b: Local industry growth and entry regressions (4-digit industries)

	(1) growth 2003-2004	(2) growth 2003-2008	(3) entry 2003-2004	(4) entry 2003-2008
log(Eirt)	-0.043***	-0.033***		
	(0.0028)	(0.0011)		
log(Eirt_SR)	0.200***	0.146***	0.051***	0.104***
	(0.0333)	(0.0120)	(0.0140)	(0.0242)
log(Eirt_CL)	-0.041	0.035	0.067	-0.195
	(0.2279)	(0.0826)	(0.2409)	(0.3628)
log(Eirt_IO)	0.042**	0.005	0.015	-0.019
	(0.0192)	(0.0077)	(0.0127)	(0.0191)
R-squared	0.082	0.201	0.153	0.207
N	21,855	21,050	9,309	9,309

Idem Table 7a, but using industries aggregated at the 4-digit level.
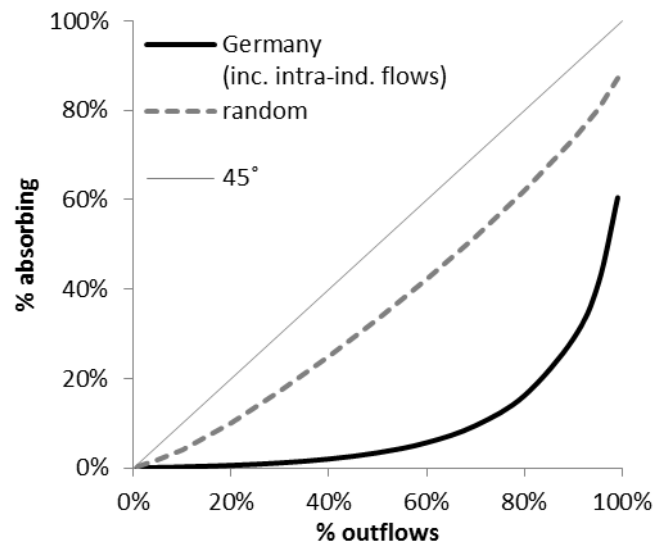
Figures

Figure 1: Flexibility of the German labor force

The vertical axis shows the percentage of employment represented by industries that together absorb the percentage of industry switchers depicted on the horizontal axis. Values include within-industry flows, are averaged across the years 1999-2007 and are calculated for the German labor market as a whole. The dotted line provides a random benchmark. The thin grey line represents the 45° as a point of reference.
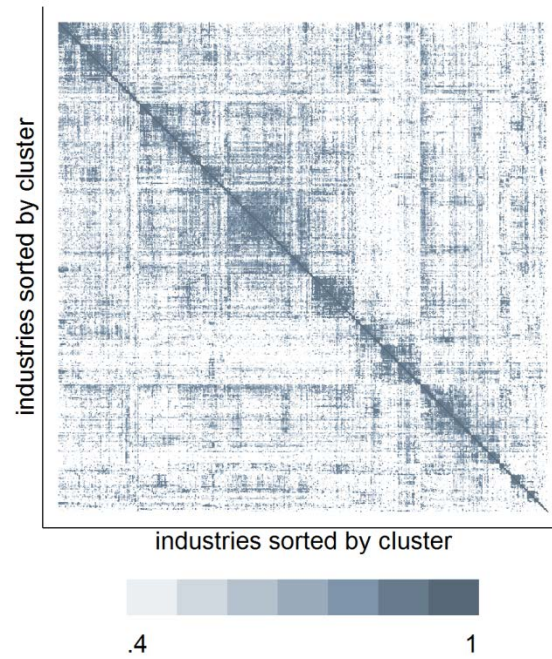
Figure 2a: Clustering of inter-industry linkages in Germany (1999-2008)

The shading in the figure represents the average symmetrized skill-relatedness matrix for Germany at the 5-digit level using all yearly flows between 1999 and 2008. Rows and columns are sorted according to an average-linkages hierarchical clustering algorithm.
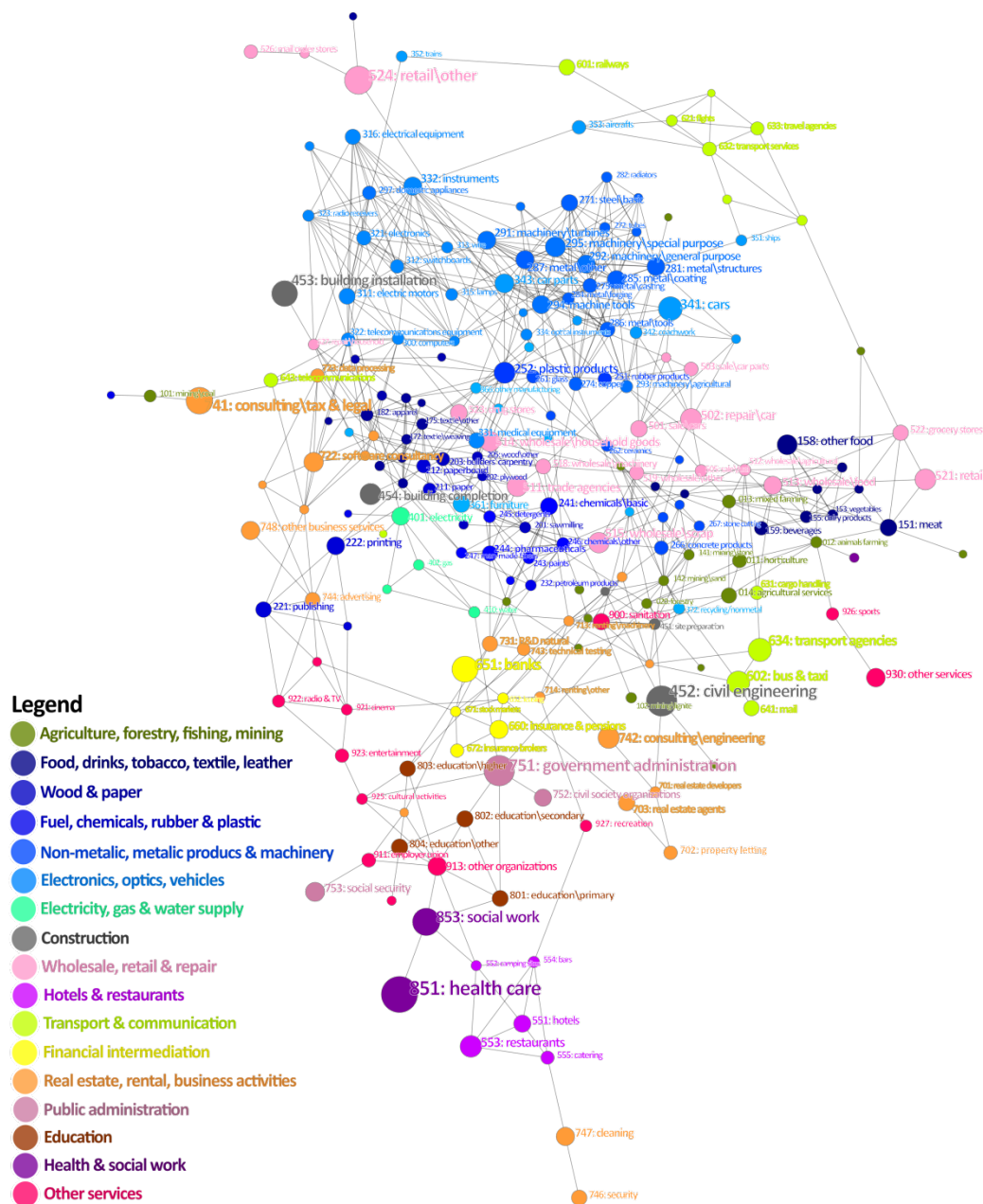
Figure 2b: Network depiction of inter-industry linkages in Germany (1999-2008)

The network depicts the strongest 651 links among 3-digit industries in the symmetrized average skill-relatedness matrix for Germany as a whole, using all yearly flows between 1999 and 2008. The layout is based on the organic layout procedure in the Cytoscape software, manually adjusted to increase the clarity of the graph by minimizing edge crossings. Labels are omitted for small industries. Color codes represent subsections in the NACE classification.
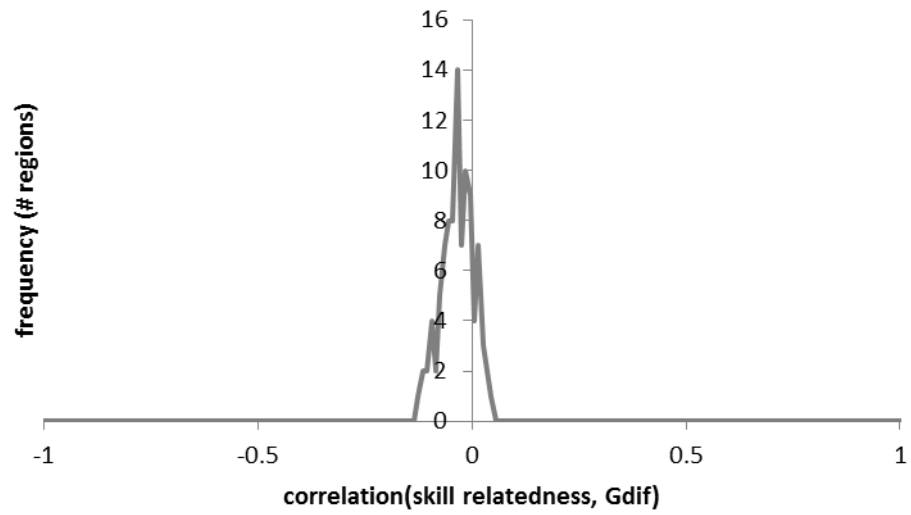
Figure 3: Regional labor reallocation: correlation($\bar{R}_{ij}$,$Gdif_{ij}$) by region

The graph shows the histogram of region-specific rank correlations between skill relatedness and the absolute difference in local growth rates within an industry pair.